Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 227–231, January 1994
Biochemistry

Crystal structure of peanut lectin, a protein with an unusual quaternary structure

("open" quaternary arrangement/protein structure/x-ray crystallography/legume lectin)

RAHUL BANERJEE, SHEKHAR C. MANDE, V. GANESH, KALYAN DAS, V. DHANARAJ, SANJEEV K. MAHANTA, K. SUGUNA, A. SUROLIA, AND M. VIJAYAN

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Communicated by David Phillips, August 27, 1993 (received for review May 15, 1993)

ABSTRACT **The x-ray crystal structure of the tetrameric T-antigen-binding lectin from peanut, M_r 110,000, has been determined by using the multiple isomorphous replacement method and refined to an *R* value of 0.218 for 22,155 reflections within the 10- to 2.95-Å resolution range. Each subunit has essentially the same characteristic tertiary fold that is found in other legume lectins. The structure, however, exhibits an unusual quaternary arrangement of subunits. Unlike other well-characterized tetrameric proteins with identical subunits, peanut lectin has neither 222 (D_2) nor fourfold (C_4) symmetry. A noncrystallographic twofold axis relates two halves of the molecule. The two monomers in each half are related by a local twofold axis. The mutual disposition of the axes is such that they do not lead to a closed point group. Furthermore, the structure of peanut lectin demonstrates that differences in subunit arrangement in legume lectins could be due to factors intrinsic to the protein molecule and, contrary to earlier suggestions, are not necessarily caused by interactions involving covalently linked sugar. The structure provides a useful framework for exploring the structural basis and the functional implications of the variability in the subunit arrangement in legume lectins despite all of them having nearly the same subunit structure, and also for investigating the general problem of "open" quaternary assembly in oligomeric proteins.**

Lectins are multivalent proteins of nonimmune origin that bind cell surface carbohydrates with high specificity (1, 2). Although originally isolated from plant sources and characterized by their ability to agglutinate erythrocytes, lectins are now known to be ubiquitous in nature, with binding specificities for a wide variety of cells. They have received considerable attention in recent years on account of their use in studies on biological receptors and cell surface phenomena.

The most extensively studied lectins are those obtained from the seeds of leguminous plants. These lectins are either dimeric or tetrameric. The first lectin to be x-ray analyzed, in the 1970s, was the tetrameric concanavalin A (Con A) from the jack bean (3, 4). The three-dimensional structures of three more lectins, those of pea lectin, favin, and isolectin I from the seeds of *Lathyrus ochrus*, became available subsequently (5–7). In the meantime, it was shown by several workers that legume lectins are related to one another by sequence homology (8). As is to be expected from the homology, the subunits in Con A, pea lectin, favin, and the *L. ochrus* lectin have nearly the same tertiary structure, although Con A has a single-chain subunit while the other three have two polypeptide chains in each subunit. They contain two metal ions each (calcium and manganese), which are situated in the same locations in the three-dimensional structures of the four lectins. The locations of the carbohydrate-binding region in them are also broadly similar. Furthermore, the lectin subunits dimerize in a similar fashion. The dimers further associate into tetramers in Con A, but not in the other three. Modes of dimerization different from those observed in the above lectins were, however, found in lectin IV of *Griffonia simplicifolia* (GS4) (9) and the *Erythrina corallodendron* lectin (10), although the structure of subunits in them is similar to that of the other legume lectins. Interactions involving covalently bound sugar were suggested to be responsible for the different modes of subunit association found in them.

The lectin from peanut (*Arachis hypogaea*), with specificity for the tumor-associated T-antigenic disaccharide Gal(β1-3)GalNAc, is, like Con A, a tetrameric protein with M_r 110,000 (11). Each subunit in the protein is 236 amino acid residues long and is homologous to the subunits in other legume lectins (12). As in the case of Con A (13, 14) and other well-known tetrameric proteins such as hemoglobin (15, 16), the molecule dissociates at low pH into dimers which bind sugar with the same stoichiometry as the tetramer (one binding site per protomer) but with an association constant one order of magnitude lower than that for the tetramer (17, 18). At physiological pH, however, the molecule is entirely tetrameric, with no evidence of association–dissociation (17). The lectin crystallizes in four forms, one orthorhombic, two monoclinic, and one triclinic, each containing a tetrameric molecule in the asymmetric unit (19, 20). Of these, the orthorhombic form was taken up for detailed study (21, 22). Attempts to solve the structure by employing the molecular replacement method using search models derived from other legume lectins did not succeed. Here we report the salient features of the structure determined subsequently by the multiple isomorphous replacement method and refined by using 2.95-Å resolution x-ray data.* These features are of considerable general interest in relation to quaternary association in proteins.

METHODS

Crystallization and Preparation of Heavy Atom Derivatives. The lectin was purified from the locally available peanut by using affinity chromatography on cross-linked arabinogalactan (23). The orthorhombic crystals ($P2_12_12$, a = 129.3 Å, b = 126.9 Å, c = 76.9 Å; one tetrameric molecule in the asymmetric unit) were grown from 0.6% protein in 0.05 M sodium phosphate buffer (pH 7.0) containing 0.2 M NaCl, 1.5 mM lactose, 0.02% sodium azide, and 12% (wt/vol) polyeth-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: GS4, lectin IV of *Griffonia simplicifolia*.

*The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1PEL, 1PELSF). This information is embargoed for 1 year (coordinates) and 2 years (structure factors) from the date of publication.

Table 1. Summary of preparation of heavy atom derivatives, data collection, and heavy atom analysis

Parameter	Native	$SmNO_3$	K_2PtCl_6	K_2PtCl_6	K_2PtCl_4	$KAuCl_4$	Iodophenyl galactoside
Heavy atom concentration, mM		3	10	7	3.25	0.85	
Duration of soaking, days		2	7	3	1	½	
Resolution, Å	2.95	3.5	3.8	3.3	3.3	3.4	3.3
No. of unique relections	23,994	16,045	11,633	19,115	19,240	17,469	19,546
Completion, %	88	97	89	98	98	97	99
R_{merge}, %	6.8	5.1	6.2	4.8	4.9	6.1	5.4
$R_{isomorphous,}$, %		21.5	32.4	24.6	21.6	34.4	17.3
Overall phasing power		0.90	0.89	0.68	0.54	1.0	0.48
No. of heavy atom sites		4	4	3	3	4	4

$R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I = observed intensity and $\langle I\rangle$ = average intensity obtained from multiple observations of symmetry-related reflections. $R_{isomorphous} = \Sigma|I_{PH} - I_P|/\Sigma I_P$, where I_P and I_{PH} are the intensities of the protein and the heavy atom derivative, respectively. Phasing power = average heavy atom structure factor divided by rms lack-of-closure.

ylene glycol (PEG) 8000 (19). An iodine derivative was prepared by replacing lactose by iodophenyl galactoside in crystallization experiments. The remaining heavy atom derivatives were obtained by controlled soaking. The derivatives included two containing K_2PtCl_6, one involving a higher concentration and a long soaking period and the other involving a lower concentration and short soaking period. Another platinum derivative was obtained by using K_2PtCl_4. A samarium derivative was prepared by using $SmNO_3$. A water solution, instead of a solution in phosphate buffer, of the components of the original mother liquor was used in soaking experiments because samarium precipitated in the presence of phosphate. A sixth derivative was obtained by using $KAuCl_4$. The compound precipitates in the presence of PEG 8000, which was therefore replaced by ammonium sulfate (25% saturation) in soaking experiments using glutaraldehyde cross-linked crystals. A summary of the conditions of soaking used in the preparation of the derivatives is given in Table 1.

Data Collection. Diffraction data from the native and derivative crystals were collected on a Siemens–Nicolet area detector mounted on a GX 20 Marconi Avionics rotating anode x-ray generator. One crystal was used for collecting each data set. The raw data were processed by using XENGEN (24). Further statistics pertaining to the data sets are given in Table 1.

Determination and Refinement of Heavy Atom Parameters. The heavy atom positions were determined by using difference Patterson and Fourier maps and the direct methods program MULTAN (25). Refinement of heavy atom parameters and phase-angle calculations were carried out using PHARE in the CCP4 program package (Daresbury Laboratory, Warrington, England). As can be seen from the relevant information given in Table 1, the phasing power of the derivatives was rather low, presumably on account of the low occupancies of the heavy atom sites. Furthermore, the three platinum derivatives have two sites in common and the two derivatives involving K_2PtCl_6 have three sites in common. Thus, the three platinum derivatives were not truly independent of one another. Therefore, the other two platinum derivatives were also excluded from phase-angle calculations when refining the heavy atom parameters in each platinum derivative.

Structure Solution and Refinement. The final phase-angle calculations yielded a mean figure of merit of 0.49 for 17,068 reflections up to a resolution of 3.3 Å. Despite the comparatively low figure of merit, a substantial part of the resulting

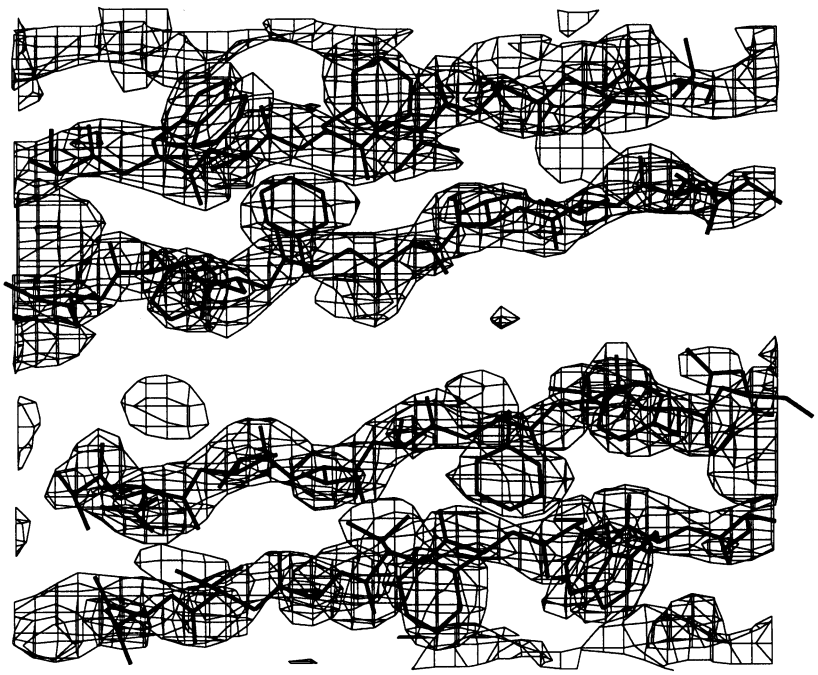

FIG. 1. A region of the electron-density map with $2|F_o| - |F_c|$ and the calculated phase angles as coefficients computed by using TURBO-FRODO (ref. 22; Biographics, Marseilles). Two strands from the flat β-sheet of subunit 1 and two from that of subunit 2 in the interface between them are illustrated. The contour level is 1.6σ.

FIG. 2. Ribbon drawing (30) of a subunit in peanut lectin. The dark and lightly shaded circles represent manganese and calcium ions, respectively.

map was readily interpretable. Detailed search revealed the presence of a molecular dyad about which the map was averaged, using Bricogne's program suite (26) in CCP4. A modified version of the molecular envelope obtained by using Wang's solvent-flattening program (27) was used for this purpose. About 82% of the atoms in two subunits, encompassing almost all the sheets and parts of the loops, could be fitted into the averaged map by using FRODO (28) on an IRIS-4D workstation, following a chain tracing made in a minimap. The two subunits and the two others related by the molecular dyad were subsequently fitted into the unaveraged map. Several cycles of refinement using the Hendrickson–Konnert restrained least squares program (29) in CCP4, model building using electron-density maps calculated with $2|F_o| - |F_c|$ as coefficients, and the gradual introduction of higher-resolution data resulted in the current model, which contains 6919 protein atoms accounting for 97% of a total of 7120 nonhydrogen atoms in the molecule. Positional and individual isotropic thermal parameters were refined. Five C-terminal residues in all four subunits and a few side-chain atoms could not be located. Solvent atoms have not yet been included in the refinement. The current R value is 0.218 for 22,155 reflections in the 10- to 2.95-Å resolution shell. The rms deviation from ideal values in bond lengths is 0.025 Å. The electron-density map corresponding to the current model, a portion of which is shown in Fig. 1, is of good quality.

RESULTS AND DISCUSSION

Structure of the Subunit. As in other legume lectins (3–10), the tertiary structure of each subunit in peanut lectin essentially consists of a flat six-stranded β-sheet and a curved seven-stranded β-sheet interconnected by loops of various lengths (Fig. 2). The two metal ions in the structure are located at the same positions as in other legume lectins. Model building using the known positions of iodine in the iodophenyl galactoside derivative and the presence of significant though diffuse electron density, presumably corresponding to lactose, at the expected locations suggest that the same is true about the carbohydrate-binding region. The four crystallographically independent subunits have nearly identical structures, the rms difference in α-carbon positions within different pairs of subunits varying between 0.6 and 0.8 Å. The differences between the tertiary structure of peanut lectin and the structures of other legume lectins are largely confined to loops and chain terminii.

Quaternary Structure. Detailed crystal structures of a number of tetrameric proteins containing identical subunits are available (31). As expected from theoretical considerations (32, 33), all of them have either 222 (D_2) symmetry, which is common, or fourfold (C_4) symmetry, which is rare. As illustrated in Fig. 3 and explained in the figure legend, the peanut lectin tetramer has neither symmetry, and the arrangement of subunits in it is different from any observed so far. Indeed, the tetrameric assembly has an "open" structure (33). As outlined later, the molecule contains three distinct intersubunit "binding sites" (33). One of them is vacant in two subunits, while another is vacant in the remaining two. Unlike in hexokinase (34), simple steric considerations do not provide a ready explanation for the absence of polymerization using these vacant sites.

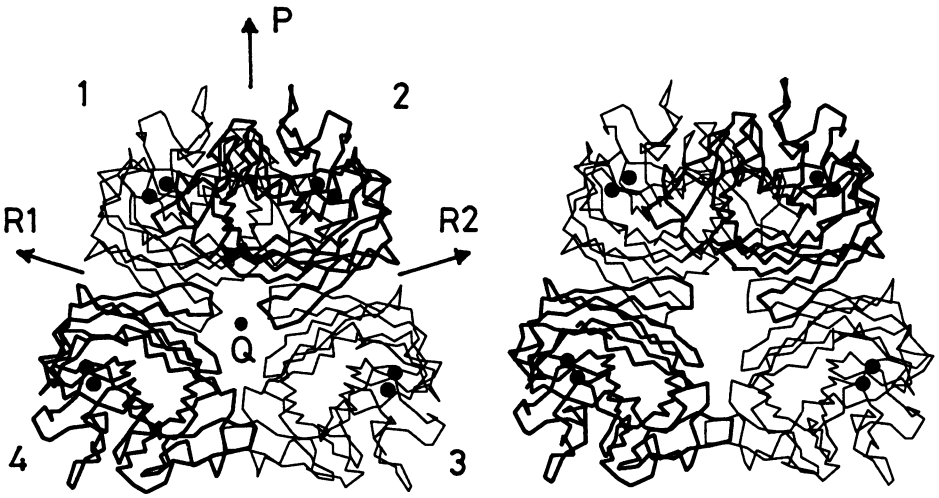

FIG. 3. A stereo view of the α-carbon backbone representation of the arrangement of subunits in the tetramer. Subunits 1 and 3 are represented by thin lines and subunits 2 and 4, by thick lines. The dots within each subunit represent metal ions. A molecular twofold axis (P) relates subunits 1 and 2 and subunits 3 and 4. Subunits 1 and 4 are related by a local twofold axis ($R1$), while 2 and 3 are related by yet another twofold axis ($R2$). $R1$ and $R2$ are inclined at 73° and −73°, respectively, with respect to P and are skewed by 12 Å on either side of P. Subunits 3 and 1 are related by a rotation of 146° about an axis (Q) passing through and perpendicular to P, $R1$, and $R2$, and a translation of 24 Å along it. The same irrational screw along Q, but in the opposite direction, relates subunits 4 and 2. This figure and Fig. 4 have been prepared by using TURBO-FRODO.

Although earlier studies (11, 17, 18) had unambiguously established peanut lectin to be a tetramer at physiological pH, it was important, in view of the unexpected results outlined above, to reconfirm that this was indeed the case. Therefore, gel filtration experiments were carried out at pH 7 at various concentrations of the protein (10, 6, and 3 mg in 1 ml) in the presence as well as the absence of lactose; in another set of experiments, 2 to 5 μg of ^{125}I-labeled protein was used in gel filtration (A.S., unpublished results). All the experiments indicated the presence of the tetramer only at the physiological pH; there was no evidence of dimers even at the very low concentrations used in the experiments involving radioactive iodine. Thus the tetrameric association in peanut lectin appears to be more stable than that in other tetrameric proteins such as Con A and hemoglobin, which exhibit tetramer–dimer equilibrium even at physiological pH (13, 14, 35). The gel filtration experiments yielded the same results irrespective of the presence of sugar. Indeed, the sugar-binding site in legume lectins, including peanut lectin, is far removed from intersubunit interfaces, and carbohydrate ligands cannot therefore physically interfere with subunit association.

The tetrameric molecule contains four subunit–subunit interfaces—namely, 1–2, 1–4, 2–3, and 3–4 (Fig. 3)—all of which are isologous. Of these 1–4 and 2–3 are related by noncrystallographic twofold symmetry. It had been shown earlier that the peanut lectin tetramer is a dimer of a dimer (21) although, contrary to expectation, the molecule does not possess 222 symmetry. Close examination suggests that subunits 1 and 4 and subunits 2 and 3 constitute the two dimers in the tetramer. The monomer–monomer interfaces in the two sets are of the same type. The other option involving the choice of subunits 1 and 2 and subunits 3 and 4 as the two dimers would lead to two types of dimers in the same molecule, which appears unlikely. Furthermore, 1–4 and 2–3 interfaces contain the largest number of intersubunit contacts, with nearly 90 interatomic distances of <4 Å involving 23 residues in each interface. The 1–2 and 3–4 interfaces contain 55 contacts involving 14 residues and 73 contacts involving 15 residues, respectively. Thus the crystal struc-

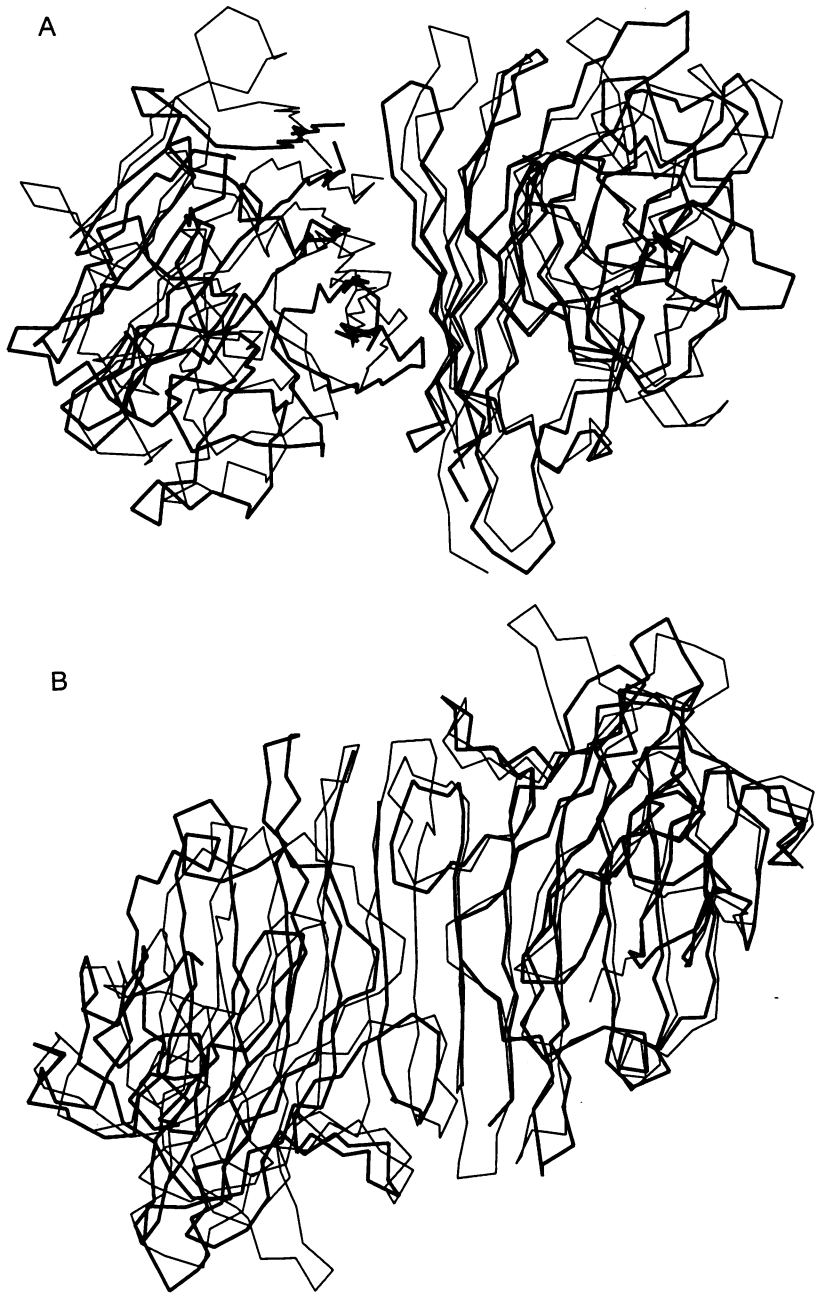

FIG. 4. α-Carbon backbone superpositions of subunits 1 and 4 and the GS4 dimer (*A*), and subunits 1 and 2 and the Con A dimer (*B*). The chains in peanut lectin are represented by thick lines and those in the other two lectins by thin lines. In both cases, subunit 1, shown on the right in *A* as well as *B*, was superposed on one of the monomers in the dimer by using the procedure of Rossmann and Argos (36). The rms deviation in α-carbon positions between subunit 1 and the monomer on which it is superposed is 1.9 Å in both *A* and *B*. The corresponding deviations between the other subunit and the other monomer are 4.8 Å and 5.1 Å, respectively. In the illustration of the "back-to-back" association involving the flat β-sheets between subunits 1 and 4 (and the two monomers in GS4) in *A*, the strands in the sheet in 1 run vertical in the plane of the figure while those in 4 run nearly perpendicular to the plane. The flat sheets are in the plane in *B*. The two sheets together form a contiguous 12-stranded β-sheet in Con A. The sheets in subunits 1 and 2 of peanut lectin do not join.

ture suggests that peanut lectin contains two identical dimers (1–4 and 2–3), both twofold symmetric, related to each other by another twofold axis in an "open" tetramer.

Relation with Subunit Arrangement in Other Legume Lectins. Dimeric association involves the formation of a 12-stranded β-sheet, six strands (of the flat β-sheet) from each subunit, in Con A (3, 4), pea lectin (5), favin (6), and isolectin I from *L. ochrus* (7). Interactions involving covalently linked sugar have been suggested to be responsible for the different modes of association found in GS4 (9) and the lectin from *E. corallodendron* (10). Peanut lectin is not a glycoprotein. Yet the most prominent intersubunit interface in it, which occurs twice (1–4 and 2–3), is remarkably similar to the monomer–monomer interface in the GS4 dimer, as can be seen from Fig. 4*A*. This interface involves the "back-to-back" association of the subunits with the two flat β-sheets, rotated with respect to each other by about 90° around an axis perpendicular to them, in contact. The 1–2 interface is similar to that in Con A dimer (Fig. 4*B*). However, the most notable feature of the latter, namely, the 12-stranded β-sheet, does not exist in peanut lectin. The flat 6-stranded β-sheets in the two subunits are farther apart in peanut lectin than in Con A, and they do not link to form the combined 12-stranded sheet. The 3–4 interface has so far not been observed in other lectins. Thus, contrary to earlier suggestions, the structure of peanut lectin, especially the occurrence of a dimeric association similar to that in GS4, demonstrates that the differences in the quaternary arrangement in legume lectins are not necessarily caused by interactions involving covalently linked carbohydrate. The tertiary structure of legume lectins appears to be such that small alterations in it could lead to different types of quaternary association.

In conclusion, the crystal structure of peanut lectin is, to the best of our knowledge, the first well-characterized example of a homotetrameric protein molecule with neither 222 nor fourfold symmetry, and it provides a framework for exploring the general problem of "open" quaternary arrangement in oligomeric proteins. It establishes that legume lectins are a class of proteins in which small alterations in essentially the same subunit structure can lead to different quaternary associations, with or without interactions involving covalently bound sugar. With three different types of intersubunit interfaces, which have various degrees of similarity with those in other legume lectins, the peanut lectin molecule provides a good model for exploring the structural basis and the functional implications of this variability.

We thank V. Sasisekharan for discussions, M. R. N. Murthy for help with computation, and C. Govindaswamy, James Paul, Stephen Suresh, K. B. Shobana, and H. K. Aruna for technical assistance. The work was supported by the Department of Science and Technology, India. The Department of Science and Technology-supported Area Detector Facility, and Department of Biotechnology-supported Interactive Graphics Facility, and the Supercomputer Education and Research Centre at the Indian Institute of Science were extensively used for data collection, model fitting, and computations. R.B. and S.K.M. have been recipients of fellowships from the University Grants Commission, while V.G. and K.D. received fellowships from the Department of Biotechnology Postdoctoral Programme and the Indian Institute of Science, respectively.

1. Liener, I. E., Sharon, D. & Goldstein, I. J., eds. (1986) in *The Lectins: Properties, Functions and Applications in Biology and Medicine* (Academic, London).
2. Sharon, N. & Lis, H. (1989) *Lectins* (Chapman & Hall, London).
3. Hardman, K. D. & Ainsworth, C. F. (1972) *Biochemistry* **11**, 4910–4919.
4. Becker, J. W., Reeke, G. N., Jr., Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975) *J. Biol. Chem.* **250**, 1513–1524.
5. Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E. & Suddath, F. L. (1986) *J. Biol. Chem.* **261**, 16518–16527.
6. Reeke, G. N., Jr., & Becker, J. W. (1986) *Science* **234**, 1108–1111.
7. Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rouge, P. & Fontecilla-Camps, J.-C. (1990) *J. Mol. Biol.* **214**, 571–584.
8. Hemperly, J. J. & Cunningham, B. A. (1983) *Trends Biochem. Sci.* **8**, 100–102.
9. Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Pearlstone, J. R., Carpenter, M. R., Smillie, L. B., Nikrad, P. V., Spohr, U. & Lemieux, R. U. (1990) *Can. J. Chem.* **68**, 1116–1121.
10. Shaanan, B., Lis, H. & Sharon, N. (1991) *Science* **254**, 862–866.
11. Lotan, R., Skutelsky, E., Danon, D. & Sharon, N. (1975) *J. Biol. Chem.* **250**, 8518–8523.
12. Young, N. M., Johnston, R. A. Z. & Watson, D. C. (1991) *Eur. J. Biochem.* **196**, 631–637.
13. Senear, D. F. & Teller, D. C. (1981) *Biochemistry* **20**, 3076–3083.
14. Senear, D. F. & Teller, D. C. (1981) *Biochemistry* **20**, 3083–3091.
15. Fanelli, A. R., Antonini, E. & Caputo, A. (1964) *Adv. Protein Chem.* **19**, 73–222.
16. Wyman, J., Jr. (1964) *Adv. Protein Chem.* **19**, 223–286.
17. Fish, W. W., Hamlin, L. M. & Miller, R. L. (1978) *Arch. Biochem. Biophys.* **190**, 693–698.
18. Decastel, M., De Boeck, H., Goussault, Y., De Bruyne, C. K., Loontiens, F. & Frenoy, J.-P. (1985) *Arch. Biochem. Biophys.* **240**, 811–819.
19. Salunke, D. M., Islam Khan, M., Surolia, A. & Vijayan, M. (1982) *J. Mol. Biol.* **154**, 177–178.
20. Salunke, D. M., Islam Khan, M., Surolia, A. & Vijayan, M. (1983) *FEBS Lett.* **156**, 127–129.
21. Salunke, D. M., Swamy, M. J., Islam Khan, M., Mande, S. C., Surolia, A. & Vijayan, M. (1985) *J. Biol. Chem.* **260**, 13576–13579.
22. Mande, S. C., Raghunathan, S., Salunke, D. M., Islam Khan, M., Swamy, M. J., Surolia, A. & Vijayan, M. (1988) *Ind. J. Biochem. Biophys.* **25**, 166–171.
23. Mazumdar, T. & Surolia, A. (1978) *Prep. Biochem.* **8**, 119–131.
24. Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. & Salamme, F. R. (1987) *J. Appl. Crystallogr.* **20**, 383–387.
25. Main, P., Fiske, S. J., Hull, S. E., Lessinger, L., Germain, G., Declercq, J.-P. & Woolfson, M. M. (1984) MULTAN 11/84, A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data (Univ. of York, England, and Louvain, Belgium).
26. Bricogne, G. (1976) *Acta Crystallogr.* **A32**, 832–847.
27. Wang, B. C. (1985) *Methods Enzymol.* **115**, 90–112.
28. Jones, T. A. (1978) *J. Appl. Crystallogr.* **11**, 268–272.
29. Hendrickson, W. A. & Konnert, J. H. (1980) in *Biomolecular Structure, Function and Evolution*, ed. Srinivasan, R. (Pergamon, Oxford, U.K.), Vol. 1, pp. 43–57.
30. Carson, M. (1987) *J. Mol. Graphics* **5**, 103–106.
31. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.
32. Monod, J., Wyman, J. & Changeux, J.-P. (1965) *J. Mol. Biol.* **12**, 88–118.
33. Cornish-Bowden, A. J. & Koshland, D. E., Jr. (1971) *J. Biol. Chem.* **246**, 3092–3102.
34. Steitz, T. A., Fletterick, R. J., Anderson, W. F. & Anderson, C. M. (1976) *J. Mol. Biol.* **104**, 197–222.
35. Ip, S. H. C. & Ackers, G. K. (1977) *J. Biol. Chem.* **252**, 82–87.
36. Rossmann, M. G. & Argos, P. (1975) *J. Biol. Chem.* **250**, 7525–7532.